Docket No.: 324-157 **PATENT**

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of	:	
CHARLET, Delphine	:	
U.S. Patent Application No. n/a	:	Group Art Unit:
Filed: Herewith	:	Examiner:

For: VERIFICATION SCORE NORMALIZATION IN A SPEAKER VOICE RECOGNITION DEVICE

CLAIM OF PRIORITY AND TRANSMITTAL OF CERTIFIED PRIORITY DOCUMENT

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

In accordance with the provisions of 35 U.S.C. 119, Applicant hereby claims the priority of French Patent Application No. 02-09299, filed July 22, 2002 in the present application. The certified copy is submitted herewith.

Respectfully submitted,

LOWE HAUPTMAN GILMAN & BERNER, LLP

Allan M. Lowe
Registration No. 19,641

BY: Michael G. Gilman
Registration No. 19,114

1700 Diagonal Road, Suite 310
Alexandria, Virginia 22314
(703) 684-1111 AML/ssw
Facsimile: (703) 518-5499
Date: July 21, 2003

RÉPUBLIQUE FRANÇAISE

02 09 99

INPI
INSTITUT NATIONAL DE LA PROPRIETE INDUSTRIELLE

BREVET D'INVENTION

CERTIFICAT D'UTILITÉ - CERTIFICAT D'ADDITION

COPIE OFFICIELLE

Le Directeur général de l'Institut national de la propriété industrielle certifie que le document ci-annexé est la copie certifiée conforme d'une demande de titre de propriété industrielle déposée à l'Institut.

Fait à Paris, le 06 JUIN 2003

Pour le Directeur général de l'Institut
national de la propriété industrielle
Le Chef du Département des brevets

Martine PLANCHE

INSTITUT NATIONAL DE LA PROPRIETE INDUSTRIELLE
SIEGE
26 bis, rue de Saint Petersbourg
75800 PARIS cedex 08
Téléphone : 33 (0)1 53 04 53 04
Télécopie : 33 (0)1 53 04 45 23

Tel dépot

INPI
INSTITUT NATIONAL DE LA PROPRIÉTÉ INDUSTRIELLE

26 bis, rue de Saint Pétersbourg
75800 Paris Cedex 08
Téléphone : 01 53 04 53 04 Télécopie : 01 42 94 86 54

BREVET D'INVENTION
CERTIFICAT D'UTILITÉ

Code de la propriété intellectuelle - Livre VI

cerfa N° 11354*01

REQUÊTE EN DÉLIVRANCE 1/2

Cet imprimé est à remplir lisiblement à l'encre noire 08 540 w /260899

Réservé à l'INPI

REMISE DES PIÈCES
DATE 22 JUIL. 2002
LIEU 99

N° D'ENREGISTREMENT NATIONAL ATTRIBUÉ PAR L'INPI 0209299

DATE DE DÉPÔT ATTRIBUÉE PAR L'INPI 22 JUIL. 2002

Vos références pour ce dossier *(facultatif)* VP/CNET04332

1 NOM ET ADRESSE DU DEMANDEUR OU DU MANDATAIRE À QUI LA CORRESPONDANCE DOIT ÊTRE ADRESSÉE

CABINET MARTINET & LAPOUX
43 Boulevard Vauban
BP 405 GUYANCOURT
78055 ST QUENTIN YVELINES CEDEX

Confirmation d'un dépôt par télécopie ☐ N° attribué par l'INPI à la télécopie

2 NATURE DE LA DEMANDE	Cochez l'une des 4 cases suivantes
Demande de brevet	☒
Demande de certificat d'utilité	☐
Demande divisionnaire	☐
Demande de brevet initiale	N° Date / /
ou demande de certificat d'utilité initiale	N° Date / /
Transformation d'une demande de brevet européen *Demande de brevet initiale*	☐ N° Date / /

3 TITRE DE L'INVENTION (200 caractères ou espaces maximum)

Normalisation de score de vérification dans un dispositif de reconnaissance vocale de locuteur

4 DÉCLARATION DE PRIORITÉ OU REQUÊTE DU BÉNÉFICE DE LA DATE DE DÉPÔT D'UNE DEMANDE ANTÉRIEURE FRANÇAISE	
	Pays ou organisation Date / / N°
	Pays ou organisation Date / / N°
	Pays ou organisation Date / / N°
	☐ S'il y a d'autres priorités, cochez la case et utilisez l'imprimé «Suite»

5 DEMANDEUR		☐ S'il y a d'autres demandeurs, cochez la case et utilisez l'imprimé «Suite»
Nom ou dénomination sociale		**FRANCE TELECOM**
Prénoms		
Forme juridique		**Société Anonyme**
N° SIREN		**380 129 866**
Code APE-NAF		
Adresse	Rue	**6, Place d'Alleray**
	Code postal et ville	**75015** **PARIS**
Pays		**FRANCE**
Nationalité		**Française**
N° de téléphone *(facultatif)*		
N° de télécopie *(facultatif)*		

INPI
INSTITUT NATIONAL DE LA PROPRIÉTÉ INDUSTRIELLE

CERTIFICAT D'UTILITÉ

REQUÊTE EN DÉLIVRANCE 2/2

Réservé à l'INPI

REMISE DES PIÈCES DATE 22 JUIL. 2002

LIEU 99

N° D'ENREGISTREMENT NATIONAL ATTRIBUÉ PAR L'INPI 0209299

DB 540 W /260899

Vos références pour ce dossier : *(facultatif)*		VP/CNET04332
6 MANDATAIRE		
Nom		LAPOUX
Prénom		Roland
Cabinet ou Société		Cabinet MARTINET & LAPOUX
N °de pouvoir permanent et/ou de lien contractuel		
Adresse	Rue	43 Boulevard Vauban BP 405 GUYANCOURT
	Code postal et ville	78055 ST QUENTIN YVELINES CEDEX
N° de téléphone *(facultatif)*		01.30.64.90.09
N° de télécopie *(facultatif)*		01.30.64.90.02
Adresse électronique *(facultatif)*		Martinet@wanadoo.fr
7 INVENTEUR (S)		
Les inventeurs sont les demandeurs		☐ Oui ☒ Non **Dans ce cas fournir une désignation d'inventeur(s) séparée**
8 RAPPORT DE RECHERCHE		**Uniquement pour une demande de brevet (y compris division et transformation)**
Établissement immédiat ou établissement différé		☒ ☐
Paiement échelonné de la redevance		**Paiement en trois versements, uniquement pour les personnes physiques** ☐ Oui ☐ Non
9 RÉDUCTION DU TAUX DES REDEVANCES		**Uniquement pour les personnes physiques** ☐ Requise pour la première fois pour cette invention *(joindre un avis de non-imposition)* ☐ Requise antérieurement à ce dépôt *(joindre une copie de la décision d'admission pour cette invention ou indiquer sa référence)* :
Si vous avez utilisé l'imprimé «Suite», indiquez le nombre de pages jointes		

10 SIGNATURE DU DEMANDEUR OU DU MANDATAIRE (Nom et qualité du signataire)

Roland LAPOUX
Mandataire
(CPI-92-1136)

VISA DE LA PRÉFECTURE OU DE L'INPI

La loi n°78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés s'applique aux réponses faites à ce formulaire. Elle garantit un droit d'accès et de rectification pour les données vous concernant auprès de l'INPI.

Normalisation de score de vérification dans un dispositif de reconnaissance vocale de locuteur

La présente invention concerne la reconnaissance vocale automatique de locuteur, et plus particulièrement la vérification d'un locuteur autorisé pour accéder à une application de service, indépendamment, ou bien en dépendance du contenu du segment vocal, tel que mot de passe, que prononce le locuteur.

La vérification du locuteur, ou encore authentification vocale, constitue un mode ergonomique pour la sécurisation d'accès. Malheureusement, ses performances actuelles n'assurent pas une sécurité totale.

Un développeur de moyen de vérification de locuteur dans un dispositif de reconnaissance automatique de parole, objet de l'invention, doit faire un compromis entre un taux de fraude autorisée correspondant à des imposteurs accédant à l'application et le niveau d'ergonomie requis correspondant à un taux d'acceptation de locuteurs de bonne foi auxquels l'application de service ne peut être refusée.

Le compromis entre sécurité et ergonomie conditionne la valeur d'un seuil de décision. En effet, tout procédé de vérification de locuteur aboutit à un score de vérification qui traduit la similarité entre un modèle vocal de locuteur autorisé présumé et un segment vocal de locuteur inconnu souhaitant accéder à l'application. Le score de vérification est ensuite comparé au seuil de décision. Selon le résultat de cette comparaison, le dispositif décide d'accepter ou de rejeter le

locuteur inconnu, c'est-à-dire de l'autoriser ou l'interdire à accéder à l'application. Si le seuil de décision est sévère et donc élevé, on acceptera à tort peu d'imposteurs mais on rejettera des locuteurs autorisés. Si le seuil de décision est lâche et donc faible, on rejettera peu de locuteurs autorisés mais on acceptera beaucoup d'imposteurs.

La difficulté réside donc dans la détermination du seuil de décision, d'autant que pour un même taux d'acceptation, le seuil est variable d'un locuteur à l'autre ("A COMPARISON OF A PRIORI THRESHOLD SETTING PROCEDURES FOR SPEAKER VERIFICATION IN THE CAVE PROJECT", J.-B. PIERROT et al., Proceedings ICASSP, 1998).

Ainsi la distribution des scores de vérification dépend du modèle vocal de locuteur sur lesquels ils sont calculés. Un fonctionnement optimal de la vérification de locuteur nécessite donc un seuil de décision respectif par modèle.

Une façon de s'affranchir de la sensibilité au seuil par locuteur réside dans la normalisation de la distribution des scores de vérification. Si par une transformation appropriée, les distributions des scores sont rendues indépendantes du modèle de locuteur, on résout alors le problème de la recherche d'un seuil par locuteur, c'est-à-dire par modèle de locuteur. Le problème est donc déplacé vers la recherche d'une normalisation des scores.

Dans la méthode dite "z-norm" selon l'article intitulé "A MAP APPROACH, WITH SYNCHRONOUS DECODING AND UNIT-BASED NORMALIZATION FOR TEXT-DEPENDENT SPEAKER VERIFICATION", Johnny MARIETHOZ et al., Proceedings ICASSP, 2000, la distribution des scores de vérification est normalisée par des paramètres μ_X

et σ_X de la distribution des scores d'imposteurs estimés sur une population d'imposteurs. Si $S_X(Y)$ est le score de vérification pour un segment vocal à tester Y par rapport à un modèle de locuteur autorisé X, le score de vérification normalisé par la méthode z-norm est :

$$\tilde{s}_X(Y) = \frac{s_X(Y) - \mu_X}{\sigma_X}$$

où μ_X et σ_X sont respectivement la moyenne et l'écart-type de la distribution des scores d'imposteurs sur le modèle X. Ces paramètres de normalisation sont estimés préalablement, lors de la phase d'apprentissage du dispositif, avec une base de données d'enregistrements qui sont considérés comme des occurrences plausibles d'imposture pour le modèle de locuteur X.

La nécessaire base de données d'enregistrements de locuteurs considérés comme imposteurs par rapport au locuteur autorisé est concevable dans le cas de la vérification de locuteur en fonction d'un mot de passe fixé et connu du dispositif de reconnaissance vocale. Cela suppose que le développeur de l'application de service aura fait auparavant une collecte d'enregistrements de personnes prononçant le mot de passe dans un contexte proche de l'application pour que ces enregistrements représentent des occurrences plausibles de tests d'imposture. Cette nécessaire collecte d'enregistrements rend difficile le changement de mot de passe dans le cas d'un système à mot de passe fixé par le dispositif et rend impossible le choix du mot de passe par le locuteur autorisé, utilisateur de l'application.

En effet, dans le cas ergonomique où le mot de passe est choisi par l'utilisateur lui-même lors de

sa phase d'apprentissage, il est pratiquement impossible d'effectuer une collecte d'enregistrements de ce mot de passe par un ensemble d'autres locuteurs.

D'autre part, pour améliorer l'ergonomie de certaines applications est prévue une phase d'apprentissage, dite enrôlement, très courte au cours de laquelle une empreinte vocale du locuteur utilisateur autorisé est créée en générant un modèle vocal de celui-ci.

Pour enrichir la modélisation, le modèle vocal de locuteur autorisé est adapté au fur et à mesure des utilisations avec des enregistrements de parole validés par l'application ou par un algorithme de décision, comme divulgué par l'article "ROBUST METHODS OF UPDATING MODEL AND A PRIORI THRESHOLD IN SPEAKER VERIFICATION", Tomoko MATSUI et al., Proceedings ICASSP, 1996, p. 97-100. Lorsqu'un utilisateur a été bien reconnu, sa parole enregistrée pendant la demande d'accès est utilisée pour mettre à jour son modèle. Cette mise à jour enrichit la modélisation et prend en compte les évolutions de la voix du locuteur autorisé au cours du temps.

Puisque la modélisation s'enrichit, la distribution des scores est modifiée et le seuil de décision défini initialement peut être inadapté à l'application. En effet, plus le modèle est déterminé avec beaucoup de données, meilleurs sont les scores de vérification dans le cas d'un locuteur-utilisateur autorisé. Si le seuil de décision est positionné assez lâche pour ne pas rejeter trop d'utilisateurs autorisés dans la configuration initiale, il est également assez permissif et laisse passer un grand nombre d'imposteurs. Comme le modèle vocal de

locuteur est enrichi au fur et à mesure des demandes d'accès, les distributions des scores sont modifiées, ce qui peut conduire à un très faible rejet des locuteurs autorisés et à un taux d'acceptation des imposteurs relativement élevé, alors qu'une modification du seuil de décision bénéficierait pleinement de l'enrichissement de la modélisation et conserverait un faible rejet à tort tout en ayant un faible taux d'acceptation d'imposteurs.

Dans l'article précité, MATSUI et al. proposent d'adapter le seuil de décision lorsque le modèle de locuteur est adapté. Cette adaptation est donc faite directement sur le seuil de décision pour un point de fonctionnement attendu.

L'adaptation du seuil proposé par MATSUI et al. suppose que le dispositif a conservé tous les enregistrements de parole nécessaires à l'apprentissage et l'adaptation du modèle de locuteur pour pouvoir déterminer un ensemble de scores de vérification qui vont servir à l'estimation d'un seuil de décision pour cet ensemble. Ce seuil est interpolé avec l'ancien seuil pour obtenir le nouveau seuil.

Les inconvénients de cette adaptation de seuil sont les suivants. D'une part, des occurrences d'enregistrements d'imposteurs sont nécessaires, ce qui est irréaliste dans certaines applications. D'autre part, les enregistrements de parole de locuteur doivent être conservés pour ré-estimer le seuil de décision ce qui implique un coût en mémoire non négligeable. Enfin, la ré-estimation étant faite au niveau du seuil de décision, c'est-à-dire pour un point de fonctionnement recherché, si l'on souhaite modifier le point de fonctionnement pour des

considérations ergonomiques par exemple, alors tous les paramètres de l'interpolation sont à modifier.

L'objectif principal de l'invention est de normaliser le score de vérification pour qu'il soit comparé à un seuil de décision toujours pertinent, indépendant du locuteur, tout en assurant que le score de vérification évolue avec la voix du locuteur autorisé, sans recourir à des enregistrements supplémentaires d'imposteurs. En conséquence, relativement au dispositif de reconnaissance de parole, l'invention vise à réduire l'espace de mémoire nécessaire à des enregistrements supplémentaires d'imposteurs, tout en garantissant une décision plus précise et rapide.

Pour atteindre cet objectif, un dispositif pour reconnaître automatiquement la voix d'un locuteur autorisé à accéder à une application, comprend un moyen pour générer des paramètres d'un modèle vocal d'acceptation relatif à un segment vocal prononcé par le locuteur autorisé et des paramètres d'un modèle vocal de refus préalablement pendant une phase d'apprentissage, un moyen pour normaliser par des paramètres de normalisation un score de vérification de locuteur dépendant du rapport de vraisemblances entre un segment vocal à tester et les modèles d'acceptation et de refus, et un moyen pour comparer le score de vérification normalisé à un premier seuil afin de n'autoriser l'accès du locuteur ayant prononcé le segment vocal à tester à l'application que si le score normalisé est au moins aussi grand que le premier seuil. Ce dispositif est caractérisé, selon l'invention, en ce qu'il comprend un moyen pour mettre à jour au moins l'un des paramètres de

normalisation en fonction d'une valeur précédente dudit paramètre et du score de vérification à chaque test de segment vocal seulement lorsque le score normalisé est au moins égal à un deuxième seuil qui est au moins égal au premier seuil.

L'expression "au moins égal à" signifie une variable supérieure ou égale à un seuil.

Si l'on souhaite modifier le point de fonctionnement, le premier seuil est modifié sans nécessiter l'ajustement des paramètres.

Le score normalisé est ainsi mis à jour en ligne, au fur et à mesure des tentatives de vérification de locuteur et donc des demandes d'accès à l'application, si bien que le score normalisé évolue avec les changements de la voix du locuteur. La mise à jour en fonction au moins d'un paramètre et non d'un seuil permet de modifier le score de décision normalisé indépendamment du point de fonctionnement requis par l'application.

Le paramètre de normalisation mis à jour peut être représentatif de la valeur moyenne statistique du score de vérification de locuteur ou de l'écart-type du score de vérification de locuteur, ou bien ces deux paramètres sont mis à jour.

La mise à jour du score normalisé est encore améliorée lorsque le dispositif comprend un moyen pour mettre à jour au moins l'un des paramètres du modèle d'acceptation en fonction d'une valeur précédente dudit paramètre de modèle seulement lorsque le score normalisé est au moins égal au deuxième seuil.

D'autres caractéristiques et avantages de la présente invention apparaîtront plus clairement à la lecture de la description suivante de plusieurs

réalisations préférées de l'invention en référence aux dessins annexés correspondants dans lesquels :

- la figure 1 est un bloc-diagramme schématique d'un système de télécommunications avec un serveur contenant un dispositif de reconnaissance vocale de locuteur ;

- la figure 2 est un bloc-diagramme fonctionnel d'un moyen d'apprentissage inclus dans le dispositif ; et

- la figure 3 est un bloc-diagramme fonctionnel d'un moyen de vérification de locuteur inclus dans le dispositif.

En se référant à la figure 1, on a représenté schématiquement un contexte préféré d'utilisation du dispositif de reconnaissance vocale automatique de locuteur DR selon l'invention. Ce contexte a trait un système de télécommunications client-serveur dans lequel un terminal de locuteur TE tel qu'un poste téléphonique ou un ordinateur personnel muni d'un modem, ou un terminal mobile, tel qu'un radiotéléphone mobile est relié à un serveur vocal téléphonique interactif SV contenant le dispositif DR, à travers un réseau d'accès téléphonique ou radiotéléphonique cellulaire RA. Lorsqu'un locuteur autorisé souhaite accéder à une application de service prédéterminée AP, un mot de passe MP ou une phrase prononcé par un locuteur autorisé L devant le microphone MI du terminal TE est transmis au serveur SV en réponse à une invitation de transmettre le mot de passe au cours d'un dialogue avec le serveur vocal SV. Le dispositif DR analyse le mot de passe MP et donne accès à l'application prédéterminée AP lorsque la voix de locuteur L a été correctement reconnue. Par exemple, l'application AP offre des services

gérés dans un serveur d'application SAP relié au serveur vocal SV à travers un réseau de paquets RP, tel que le réseau internet.

Selon d'autres variantes d'application, le dispositif DR est implémenté dans un terminal, tel qu'un poste téléphonique, un ordinateur personnel, un radiotéléphone mobile, ou un assistant numérique personnel.

Comme montré aux figures 2 et 3, le dispositif de reconnaissance vocale automatique de locuteur DR selon l'invention comprend fonctionnellement un moyen d'apprentissage composé de trois modules logiciels A1, A2 et A3, et un moyen de vérification automatique de locuteur composé de six modules logiciels V1 à V6. Ils coopèrent avec une portion de mémoire non volatile dans le serveur pour mémoriser divers paramètres dont la plupart sont mis à jour, servant à des déterminations de score de vérification normalisé défini plus loin.

Le moyen d'apprentissage détermine des paramètres caractérisant principalement un modèle vocal du locuteur autorisé L à reconnaître. Il comprend un module d'acquisition de parole A1 connecté à une source acoustique, tel que le microphone MI, un module d'analyse acoustique A2 dont la sortie est bouclée sur une entrée itérative de modèles vocaux pendant une phase d'apprentissage, et un module de génération de modèle de locuteur A3.

La phase d'apprentissage automatique, dite également enrôlement, est fondée par exemple sur la modélisation statistique d'un mot de passe MP par des chaînes de Markov cachées HMM (Hidden Markov Model). On pourra se reporter au sujet des méthodes

statistiques de modélisation markovienne cachée à l'article de Lawrence R. RABINER, "A Tutorial on Hidden Markov Models and Selected Applications in speech Recognition", Proceedings of the IEEE, vol. 77, No. 2, February 1989, p. 257-286. Le mot de passe MP est prononcé devant le microphone MI pendant N occurrences de parole de durée prédéterminée chacune, typiquement N = 3 fois, par le locuteur L autorisé à accéder à l'application AP dans le serveur vocal SV. N versions du mot de passe sont alors mémorisées dans le module d'acquisition A1, après conversion analogique-numérique. Le mot de passe MP est choisi librement par le locuteur L et est inconnu a priori du dispositif de reconnaissance vocale de locuteur DR. Aucun autre enregistrement du mot de passe prononcé par des locuteurs autres que le locuteur autorisé L n'est nécessaire pendant la phase d'apprentissage.

En variante, la composition des mots de passe est libre, c'est-à-dire est constituée par tout segment vocal, et peut être changée au gré du locuteur autorisé à chaque tentative de reconnaissance de sa voix.

Au fur et à mesure des versions analysées du mot de passe prononcé, le module d'analyse A2 estime, d'une manière itérative connue, des paramètres prédéterminés m d'un modèle de Markov caché λ, afin d'en déduire les moyennes de distribution gaussienne de ces paramètres de modèle. Le module A2 hérite d'autres paramètres d'un modèle vocal général qui ont été prémémorisés dans le module A2, à cause du faible nombre de données disponibles résultant de l'analyse des versions du mot de passe en petit nombre N. Les paramètres du modèle vocal d'acceptation λ ainsi

générés du locuteur autorisé L sont mémorisés dans le module A3.

Le modèle vocal λ, dit également référence acoustique, est caractéristique de la voix du locuteur autorisé L et peut être associé en mémoire du serveur SV à un identificateur du locuteur, tel qu'un code secret et composé au clavier du terminal TE avant de prononcer le mot de passe MP.

Parallèlement à la construction du modèle d'acceptation λ, le module d'analyse acoustique A2 construit un modèle vocal de refus ϖ, dit également modèle alternatif (background model) ou anti-model. Les paramètres du modèle de refus ϖ sont connus et pré-mémorisés dans le serveur SV pendant la phase d'apprentissage. Ils sont représentatifs d'un modèle vocal "moyen" d'un nombre élevé de locuteurs quelconques, et par conséquent d'un modèle vocal d'imposture.

A la fin de la phase d'apprentissage, le module de génération A3 détermine des valeurs initiales de paramètres $\tilde{\mu}_{\lambda 0}$ et $\tilde{\tau}_{\lambda 0}$ nécessaires à la normalisation de score de vérification selon l'invention, estimées sur un corpus de données d'apprentissage définies préalablement notamment en fonction de l'application AP à laquelle le locuteur accède par le mot de passe reconnu. Ces données d'apprentissage ont été écrites préalablement dans la mémoire du serveur SV et permettent au module A3 de déterminer des valeurs initiales $\tilde{\mu}_{\lambda 0}$ et $\tilde{\tau}_{\lambda 0}$ de paramètres de normalisation dépendant notamment de paramètres des modèles vocaux λ et ϖ et utilisées dans des formules récurrentes de ces paramètres lors d'un premier test, et des facteurs d'adaptation τ_μ et τ_σ respectivement pour les paramètres de normalisation $\tilde{\mu}_\lambda$ et $\tilde{\sigma}_\lambda$.

En variante, au lieu de générer des modèles paramétriques du type HMM, les modèles d'acceptation et de refus ϖ sont générés selon une modèlisation GMM (Gaussian Mixture Model) fondée sur le mélange de distributions normales, dites distributions gaussiennes, relatives à des paramètres. La modélisation GMM est par exemple définie dans l'article de Douglas A. REYNOLDS, "Speaker identification and verification using Gaussian mixture speaker models", Speech Communication 17, 1995, p. 91-108.

Lors d'une tentative d'accès à l'application AP, par exemple après une validation du code secret composé précité par le serveur vocal SV, le locuteur L prononce devant le microphone MI un segment vocal contenant le mot de passe MP, soit une occurrence de signal de parole X pendant une durée T, afin que la chaîne des modules V1 à V6 montrée à la figure 3 vérifie que le locuteur est bien celui qui a prononcé le mot de passe pendant la phase d'apprentissage. La durée T est exprimée en nombre de portions de durée prédéterminée de 32 ms environ du segment vocal, appelées trames (frames). Le nombre T est variable en fonction de la vitesse de locution du locuteur.

Les modules d'acquisition A1 et A2 analysent acoustiquement le signal X contenant le mot de passe MP qui vient d'être prononcé, et produisent un signal vocal de test X composé d'une suite de T vecteurs de coefficients cepstraux.

Des modules de similarité V1 et V2 évaluent les similarités entre le signal vocal de test X produit par le module d'analyse acoustique A2 d'une part, et le modèle vocal d'acceptation λ et le modèle vocal de refus ϖ lus en mémoire par le module A3 d'autre

part, les paramètres m des modèles λ et ϖ ayant été mis à jour à la fin de la vérification de voix de locuteur précédente, comme on le verra plus loin. Les similarités sont exprimées par des probabilités conditionnelles $P(X|\lambda)$ et $P(X|\varpi)$ respectivement produites par les modules V1 et V2 et caractérisant la vraisemblance que le signal vocal de test observé X soit représentatif du locuteur autorisé ayant prononcé un segment vocal représenté par le modèle d'acceptation λ et la vraisemblance que le signal vocal de test observé X soit représentatif de n'importe quel locuteur ayant pu prononcé un segment vocal représenté par le modèle de refus ϖ.

Le module V3 détermine le score de vérification S_V en fonction des probabilités produites, selon la relation suivante :

$$S_V = \frac{1}{T}(\log P(X / \lambda) - \log P(X / \varpi)).$$

Le score est proportionnel au rapport de vraisemblances relatives au modèle d'acceptation λ représentatif du locuteur autorisé et au modèle de refus ϖ représentatif de n'importe quel locuteur. Il exprime la confiance accordée au signal vocal de test observé X. Plus le score S_V est élevé, plus la voix du locuteur à l'origine du signal vocal de test X présente des caractéristiques proches de celles du modèle d'acceptation λ. T dénote le nombre de trames (frames) contenues dans le segment vocal MP à tester.

Le module V3 détermine également un score de vérification normalisé S_N en fonction du score de vérification de locuteur S_V et de deux paramètres de normalisation $\tilde{\mu}_\lambda$ et $\tilde{\sigma}_\lambda$ du modèle d'acceptation λ représentatif de la voix du locuteur autorisé L, selon la relation suivante :

$$S_N = \frac{S_V - \tilde{\mu}_\lambda}{\tilde{\sigma}_\lambda}.$$

Les deux paramètres $\tilde{\mu}_\lambda$ et $\tilde{\tau}_\lambda$ résultent d'une mise à jour selon les relations de récurrence suivantes, à la fin de la vérification de locuteur ayant précédé celle en cours :

$$\tilde{\mu}_\lambda \equiv (1-\tau_\mu)\tilde{\mu}_\lambda + \tau_\mu . S_V$$

$$\tilde{\sigma}_\lambda \equiv \sqrt{(1-\tau_\sigma)\tilde{\sigma}_\lambda^2 + \tau_\sigma (S_V - \tilde{\mu}_\lambda)^2} \ .$$

Le premier paramètre de normalisation $\tilde{\mu}_\lambda$ représente la valeur moyenne statistique, c'est-à-dire l'espérance mathématique du score de vérification de locuteur. La mise à jour du premier paramètre est pondérée par un facteur d'adaptation prédéterminé τ_μ inférieur à 1. Le deuxième paramètre de normalisation $\tilde{\sigma}_\lambda$ représente l'écart-type du score de vérification S_V égal à la racine carrée de la différence de la valeur quadratique moyenne du score S_V et du carré de la valeur moyenne statistique μ_λ^2. La mise à jour du deuxième paramètre est pondérée par un autre facteur d'adaptation prédéterminé τ_σ inférieur à 1. Ainsi les paramètres de normalisation $\tilde{\mu}_\lambda$ et $\tilde{\sigma}_\lambda$ sont mis à jour en ligne par estimation de leurs moyennes sur les vérifications de locuteur précédentes.

Les valeurs des paramètres $\tilde{\mu}_\lambda$ et $\tilde{\sigma}_\lambda$ dans les membres droits des deux relations de récurrence précédentes ont été déterminées au cours de la vérification de locuteur précédant celle en cours et sont lues avec les facteurs d'adaptation τ_μ et τ_σ par le module V3 avant la détermination du score S_N. Lors de la première vérification de locuteur succédant à la phase d'apprentissage, les paramètres initiaux $\tilde{\mu}_{\lambda 0}$ et $\tilde{\sigma}_{\lambda 0}$ sont lus par le module V3 en tant que

paramètres $\tilde{\mu}_\lambda$ et $\tilde{\sigma}_\lambda$ servant à la première détermination du score normalisé S_N.

La normalisation du score de vérification de locuteur S_V en le score normalisé S_N suit avantageusement les variations du score de vérification, c'est-à-dire de la voix du locuteur, représentées par les paramètres $\tilde{\mu}_\lambda$ et $\tilde{\sigma}_\lambda$. Comme on le verra ci-après, l'évolution de la voix du locuteur autorisé L est reportée dans le score normalisé S_N par une mise à jour des paramètres $\tilde{\mu}_\lambda$ et $\tilde{\sigma}_\lambda$, mais également de paramètres de modèle m ayant servis à la modélisation initiale.

Ensuite le module de décision V4 compare le score normalisé S_N à un premier seuil prédéterminé TH1. Le module V4 autorise l'accès du locuteur à l'application de service AP lorsque le score normalisé S_N est égal ou supérieur au seuil prédéterminé TH1.

Au contraire, si S_N<TH1, l'accès à l'application de service AP est refusé au locuteur. Aucune mise à jour de paramètres n'est effectuée puisque le locuteur est considéré comme un imposteur. De préférence, le serveur vocal SV invite le locuteur à prononcer quelques fois encore le mot de passe MP, par exemple trois fois.

La décision d'accès effectuée dans le module V4 dépend du seuil TH1 constant et donc indépendant du locuteur autorisé. Selon l'invention, la décision dépend plutôt du score de vérification normalisé S_N dont les paramètres tels que les facteurs τ_μ et τ_σ sont choisis une fois pour toutes en dépendance de l'ergonomie souhaitée pour accéder à l'application AP. Si le type d'application est changé, le seuil TH1 ainsi qu'un deuxième seuil TH2 peuvent être modifiés

par le gestionnaire de la nouvelle application dans le serveur SV.

Si l'accès est autorisé, le module de validation V5 compare le seuil normalisé S_N au deuxième seuil TH2 de préférence plus grand que le premier seuil TH1, bien que les seuils puissent être égaux. Le module d'adaptation V6 ne met à jour des paramètres que si le score normalisé est plus grand que le seuil TH2, c'est-à-dire lorsque par exemple la voix du locuteur autorisé a sensiblement changée, notamment à cause du vieillissement ou d'une laryngite du locuteur.

Comme déjà dit, les paramètres de normalisation $\tilde{\mu}_\lambda$ et $\tilde{\sigma}_\lambda$ sont mis à jour selon les deux relations de récurrence ci-dessus, en fonction du score de vérification S_V qui vient d'être déterminé par le module V3 et des valeurs de paramètres $\tilde{\mu}_\lambda$ et $\tilde{\sigma}_\lambda$ qui ont été déterminées lors de la vérification de locuteur précédente.

De préférence, l'un ou les deux facteurs d'adaptation τ_μ et τ_σ varient en fonction du nombre d'adaptations, c'est-à-dire du nombre de mises à jour de paramètre de normalisation réalisées dans le module V6 depuis la phase d'apprentissage afin d'adapter rapidement les paramètres de normalisation pour qu'ils convergent rapidement lors de premières adaptations, puis de moins en moins ensuite jusqu'à suspendre l'adaptation. Plus le facteur de vitesse d'adaptation τ_μ, τ_σ est grand, plus l'adaptation du paramètre $\tilde{\mu}_\lambda$, $\tilde{\sigma}_\lambda$ est rapide.

Le module V6 met également à jour chaque paramètre m au moins du modèle d'acceptation λ et éventuellement du modèle de refus ϖ, de manière à diminuer le taux d'imposture représenté par la probabilité $P(X|\varpi)$. La mise à jour de chaque

paramètre de modèle m est basée sur une adaptation incrémentable selon la relation de récurrence suivante :

$$m = \frac{N_{AP} m_{AP} + N_{adapt} m_{adapt}}{N_{AP} + N_{adapt}} .$$

m_{AP} et N_{AP} dénotent respectivement la moyenne de la distribution gaussienne, dite également distribution normale, de la densité de probabilité du paramètre de modèle m au cours de la phase d'apprentissage et le nombre de trames dans les segments vocaux, c'est-à-dire dans les mots de passe, ayant servi à estimer les moyennes des distributions gaussiennes relatives aux modèles de Markov cachés λ et ϖ. Le paramètre m_{adapt} dénote la moyenne de la distribution gaussienne de la densité de probabilité du paramètre de modèle m qui a été déterminée lors de la mise à jour qui vient d'être réalisée et donc qui reflète l'évolution du paramètre m au cours des mises à jour, après la phase d'apprentissage. N_{adapt} dénote le nombre de trames ayant servi à estimer la moyenne de la distribution gaussienne du paramètre de modèle m pour la mise à jour qui vient d'être réalisée. Le nombre de trames T du signal vocal à tester varie d'une vérification à la suivante en fonction notamment de la vitesse de locution du locuteur.

Après la mise à jour, le module V6 mémorise les nouvelles valeurs des paramètres m des modèles vocaux λ et ϖ et des paramètres de normalisation $\tilde{\mu}_\lambda$ et $\tilde{\sigma}_\lambda$ qui serviront à la détermination des scores S_V et S_N dans le module V3 lors du prochain test de voix de locuteur.

En variante, notamment afin de diminuer la durée de chaque vérification de locuteur, seulement l'un

des paramètres de normalisation $\tilde{\mu}_\lambda$ et $\tilde{\sigma}_\lambda$ est mis à jour, de préférence seulement le paramètre de valeur moyenne statistique $\tilde{\mu}_\lambda$ ce qui s'impose en attribuant la valeur nulle au facteur d'adaptation τ_σ. De même, au moins l'un ou quelques-uns des paramètres de modèle m sont seulement mis à jour, ce qui s'impose en attribuant la valeur nulle aux nombres de trames N_{adapt} pour les autres paramètres de modèle qui ne sont pas à mettre à jour.

REVENDICATIONS

1 - Dispositif pour reconnaître automatiquement la voix d'un locuteur autorisé à accéder à une application (AP), comprenant un moyen (A1, A2, A3) pour générer des paramètres (m) d'un modèle vocal d'acceptation (λ) relatif à un segment vocal (MP) prononcé par le locuteur autorisé et des paramètres (m) d'un modèle vocal de refus (ϖ) préalablement pendant une phase d'apprentissage, un moyen (V1, V2, V3) pour normaliser par des paramètres de normalisation un score de vérification de locuteur dépendant du rapport de vraisemblances entre un segment vocal à tester (X) et les modèles d'acceptation et de refus, et un moyen (V4) pour comparer le score de vérification normalisé (S_N) à un premier seuil (TH1) afin de n'autoriser l'accès du locuteur ayant prononcé le segment vocal à tester à l'application (AP) que si le score normalisé est au moins aussi grand que le premier seuil, caractérisé en ce qu'il comprend un moyen (V6) pour mettre à jour au moins l'un ($\tilde{\mu}_\lambda$) des paramètres de normalisation en fonction d'une valeur précédente dudit paramètre et du score de vérification (S_V) à chaque test de segment vocal seulement lorsque le score normalisé (S_N) est au moins égal à un deuxième seuil (TH2) qui est au moins égal au premier seuil (TH1).

2 - Dispositif conforme à la revendication 1, dans lequel le paramètre mis à jour est représentatif de la valeur moyenne statistique ($\tilde{\mu}_\lambda$) du score de vérification de locuteur (S_V).

3 - Dispositif conforme à la revendication 2, dans lequel la valeur moyenne statistique ($\tilde{\mu}_\lambda$) du

score de vérification S_V est mise à jour selon la relation suivante :

$$\tilde{\mu}_\lambda \equiv (1 - \tau_\mu)\tilde{\mu}_\lambda + \tau_\mu . S_V$$

dans laquelle τ_μ est un facteur d'adaptation prédéterminé.

4 - Dispositif conforme à la revendication 3, dans lequel le facteur d'adaptation prédéterminé τ_μ varie en fonction du nombre de mises à jour de paramètre de normalisation.

5 - Dispositif conforme à l'une quelconque des revendications 1 à 4, dans lequel le paramètre mis à jour est représentatif de l'écart-type ($\tilde{\sigma}_\lambda$) du score de vérification de locuteur (S_V).

6 - Dispositif conforme à la revendication 5, dans lequel l'écart-type $\tilde{\sigma}_\lambda$ du score de vérification S_V est mise à jour selon la relation suivante :

$$\tilde{\sigma}_\lambda \equiv \sqrt{(1-\tau_\sigma)\tilde{\sigma}_\lambda^2 + \tau_\sigma (S_V - \tilde{\mu}_\lambda)^2}$$

dans laquelle τ_σ est un facteur d'adaptation prédéterminé.

7 - Dispositif conforme à la revendication 6, dans lequel le facteur d'adaptation prédéterminé τ_σ varie en fonction du nombre de mises à jour de paramètre de normalisation.

8 - Dispositif conforme à l'une quelconque des revendications 1 à 7, comprenant un moyen (V6) pour mettre à jour au moins l'un des paramètres (m) du modèle d'acceptation (λ) en fonction d'une valeur précédente dudit paramètre de modèle seulement

lorsque le score normalisé (S_N) est au moins égal au deuxième seuil (TH2).

9 - Dispositif conforme à la revendication 8, dans lequel le paramètre de modèle m est mis à jour selon la relation suivante :

$$m = \frac{N_{AP} m_{AP} + N_{adapt} m_{adapt}}{N_{AP} + N_{adapt}}$$

dans laquelle m_{AP} et N_{AP} dénotent respectivement la moyenne de la distribution gaussienne de la densité de probabilité du paramètre de modèle (m) au cours de la phase d'apprentissage et le nombre de trames dans les segments vocaux ayant servi à estimer des moyennes de distributions gaussiennes relatives aux modèles d'acceptation (λ) et de refus (ϖ), m_{adapt} dénote la moyenne de la distribution gaussienne de la densité de probabilité du paramètre de modèle (m) déterminée lors de la mise à jour qui vient d'être réalisée, et N_{adapt} dénote le nombre de trames ayant servi à estimer la moyenne de la distribution gaussienne du paramètre de modèle (m) pour la mise à jour qui vient d'être réalisée.

10 - Dispositif conforme à l'une quelconque des revendications 1 à 9, dans lequel le score normalisé S_N est déterminé en fonction du score de vérification de locuteur S_V et de deux paramètres de normalisation mis à jour $\tilde{\mu}_\lambda$ et $\tilde{\sigma}_\lambda$, selon la relation suivante :

$$S_N = \frac{S_V - \tilde{\mu}_\lambda}{\tilde{\sigma}_\lambda},$$

les paramètres $\tilde{\mu}_\lambda$ et $\tilde{\sigma}_\lambda$ étant respectivement la valeur moyenne statistique et l'écart-type du score de vérification de locuteur.

FIG. 1

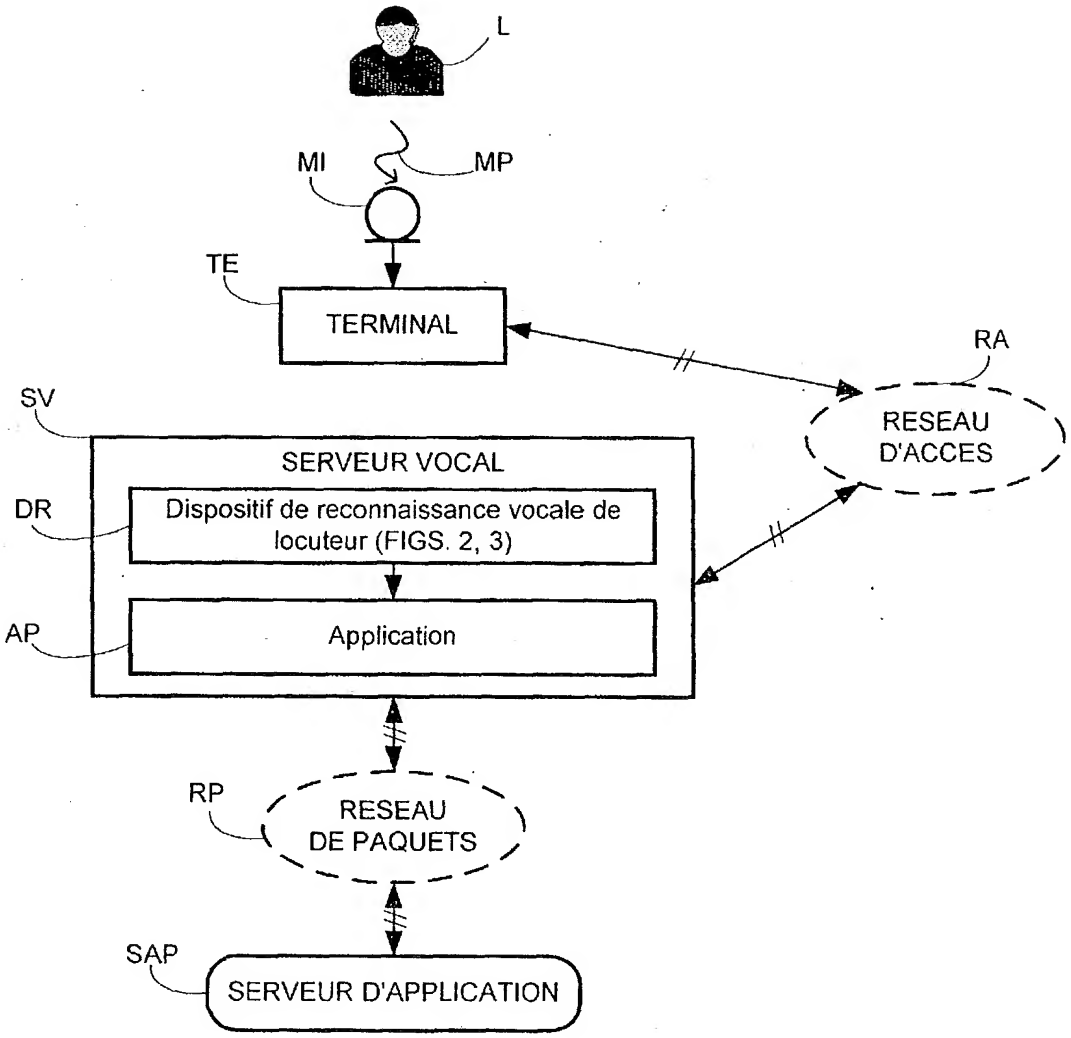

L
MI
MP
TE
TERMINAL
RA
RESEAU
D'ACCES
SV
SERVEUR VOCAL
DR
Dispositif de reconnaissance vocale de locuteur (FIGS. 2, 3)
AP
Application
RP
RESEAU
DE PAQUETS
SAP
SERVEUR D'APPLICATION

FIG. 2

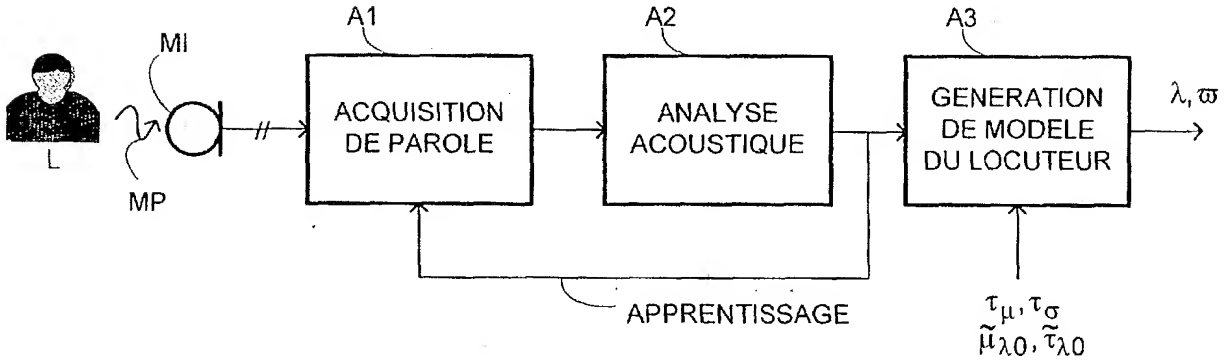
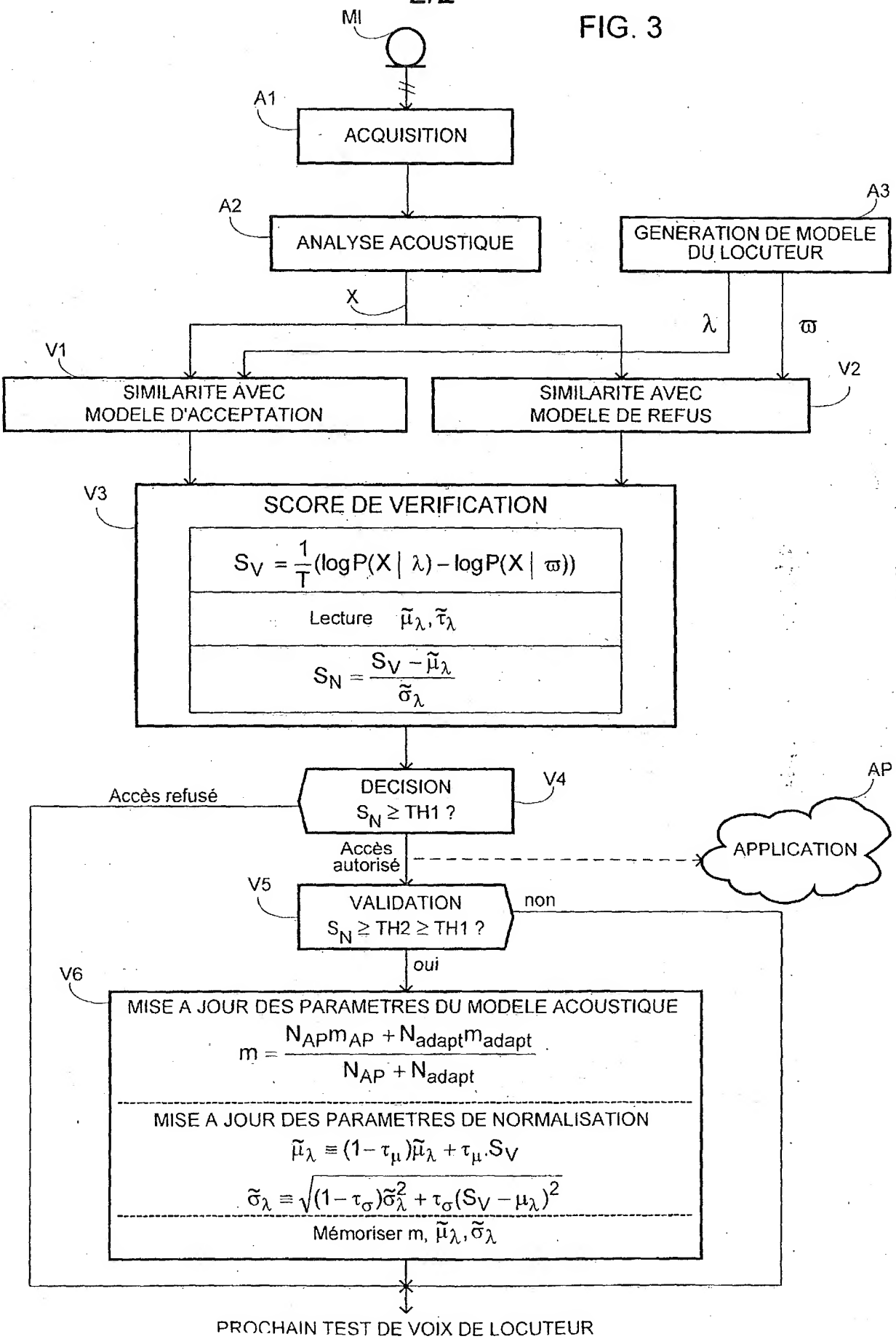

L
MI
MP
A1
ACQUISITION
DE PAROLE
A2
ANALYSE
ACOUSTIQUE
A3
GENERATION
DE MODELE
DU LOCUTEUR
λ, ϖ
APPRENTISSAGE
τμ, τσ
μ̃λ0, τ̃λ0

FIG. 3


MI
A1
ACQUISITION
A2
ANALYSE ACOUSTIQUE
A3
GENERATION DE MODELE DU LOCUTEUR
X
λ
ϖ
V1
SIMILARITE AVEC MODELE D'ACCEPTATION
V2
SIMILARITE AVEC MODELE DE REFUS
V3
SCORE DE VERIFICATION
$S_V = \frac{1}{T}(\log P(X \mid \lambda) - \log P(X \mid \varpi))$
Lecture $\tilde{\mu}_\lambda, \tilde{\tau}_\lambda$
$S_N = \frac{S_V - \tilde{\mu}_\lambda}{\tilde{\sigma}_\lambda}$
DECISION
$S_N \geq TH1$?
V4
Accès refusé
Accès autorisé
AP
APPLICATION
V5
VALIDATION
$S_N \geq TH2 \geq TH1$?
non
oui
V6
MISE A JOUR DES PARAMETRES DU MODELE ACOUSTIQUE
$m = \frac{N_{AP} m_{AP} + N_{adapt} m_{adapt}}{N_{AP} + N_{adapt}}$
MISE A JOUR DES PARAMETRES DE NORMALISATION
$\tilde{\mu}_\lambda = (1-\tau_\mu)\tilde{\mu}_\lambda + \tau_\mu . S_V$
$\tilde{\sigma}_\lambda = \sqrt{(1-\tau_\sigma)\tilde{\sigma}_\lambda^2 + \tau_\sigma (S_V - \mu_\lambda)^2}$
Mémoriser m, $\tilde{\mu}_\lambda, \tilde{\sigma}_\lambda$
PROCHAIN TEST DE VOIX DE LOCUTEUR

INPI
INSTITUT NATIONAL DE LA PROPRIETE INDUSTRIELLE

DÉPARTEMENT DES BREVETS

26 bis, rue de Saint Pétersbourg
75800 Paris Cedex 08
Téléphone : 01 53 04 53 04 Télécopie : 01 42 94 86 54

CERTIFICAT D'UTILITÉ

Code de la propriété intellectuelle - Livre VI

N° 11235*02

DÉSIGNATION D'INVENTEUR(S) Page N° .1./.1.

(Si le demandeur n'est pas l'inventeur ou l'unique inventeur)

Cet imprimé est à remplir lisiblement à l'encre noire

DB 113 W /260899

Vos références pour ce dossi r *(facultatif)*	**VP/CNET04332**
N° D'ENREGISTREMENT NATIONAL	0209299

TITRE DE L'INVENTION (200 caractères ou espaces maximum)

Normalisation de score de vérification dans un dispositif de reconnaissance vocale de locuteur

LE(S) DEMANDEUR(S) :

FRANCE TELECOM
6, Place d'Alleray
75015 PARIS

DESIGNE(NT) EN TANT QU'INVENTEUR(S) : (Indiquez en haut à droite «Page N° 1/1» S'il y a plus de trois inventeurs, utilisez un formulaire identique et numérotez chaque page en indiquant le nombre total de pages).

Nom		**CHARLET**
Prénoms		**Delphine**
Adresse	Rue	**38, rue Georges Pompidou**
	Code postal et ville	**22300** **LANNION**
Société d'appartenance *(facultatif)*		
Nom		
Prénoms		
Adresse	Rue	
	Code postal et ville	
Société d'appartenance *(facultatif)*		
Nom		
Prénoms		
Adresse	Rue	
	Code postal et ville	
Société d'appartenance *(facultatif)*		

DATE ET SIGNATURE(S) DU (DES) DEMANDEUR(S) OU DU MANDATAIRE (Nom et qualité du signataire)

Roland LAPOUX
Mandataire
(CPI/92-1136)

Le 19 Juillet 2002

La loi n°78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés s'applique aux réponses faites à ce formulaire.
Elle garantit un droit d'accès et de rectification pour les données vous concernant auprès de l'INPI ()

LAW OFFICES
LOWE HAUPTMAN GILMAN & BERNER, LLP
SUITE 300
1700 DIAGONAL ROAD
ALEXANDRIA, VIRGINIA 22314
703 — 535-7062

Atty docket #

324-157